How to get more memory for your job

Category: Memory Usage on Pleiades

DRAFT

This article is being reviewed for completeness and technical accuracy.

If your job was terminated because it needed more memory than what's available in the nodes that it ran on, consider the following:

- Among the Harpertown nodes, the 64 nodes in rack 32 have 16 GB per node instead of 8 GB per node. You can request running your job on rack 32 with the keyword **bigmem=true**. For example, change

  ```
  #PBS -lselect=1:ncpus=8
  ```

  to

  ```
  #PBS -lselect=1:ncpus=8:bigmem=true
  ```
- Run your job on Nehalem-EP or Westmere nodes instead of Harpertown nodes. For example, change

  ```
  #PBS -lselect=1:ncpus=8:model=har
  ```

  to

  ```
  #PBS -lselect=1:ncpus=8:model=neh
  ```

  or

  ```
  #PBS -lselect=1:ncpus=8:model=wes
  ```

- If all processes use about the same amount of memory and you can not fit 8 processes per node (for Harpertown or Nehalem-EP, or 12 processes per node for Westmere-EP), reduce the number of processes per node and request more nodes for your job. For example, change

  ```
  #PBS -lselect=3:ncpus=8:mpiprocs=8:model=neh
  ```

  to

```
#PBS -lselect=6:ncpus=4:mpiprocs=4:model=neh
```

- For a typical MPI job where rank 0 does the I/O and uses a lot of buffer cache, assign rank 0 to 1 node by itself. For example, change

  ```
  #PBS -lselect=1:ncpus=8:mpiprocs=8:model=neh
  ```

  to

```
#PBS
-lselect=1:ncpus=1:mpiprocs=1:model=neh+1:ncpus=7:mpiprocs=7:model=neh

Due to formatting issue, the above may appear as 2 lines. It should
really be just 1 line.
```

- If you suspect that certain nodes that your job ran on had less total physical memory than normal, report it to NAS Help Desk. Those nodes can be offlined and taken care of by NAS staff. This prevents you and other users from using those nodes before they are fixed.

- For certain pre- or post-processing work that needs more than 22.5 GB of memory, run it on the bridge nodes (bridge[1,2]) interactively. Note that jobs running on the bridge nodes can not use more than 48 GB of memory. Also MPI applications that use SGI's MPT library can not run on the bridge nodes.

- For a multi-process or multi-thread job, if any of your processes/threads needs more than 22.5 GB, it won't run on Pleiades. Run it on a shared memory system such as Columbia.

Article ID: 222
Last updated: 24 May, 2011
Computing at NAS -> Best Practices -> Memory Usage on Pleiades -> How to get more memory for your job
http://www.nas.nasa.gov/hecc/support/kb/entry/222/?ajax=1